ITEA Journal 2011; 32: 21–28

Operational Testing of Software-Intensive Systems: Observations and Comments

Will Manning
U.S. Army Operational Test Command, Fort Hood, Texas

Three operational test events from separate ACAT 1D acquisition programs are the foundation for this study. The generalized effect of the software systems on individual test results is examined. The test results support discussions related to several current test and evaluation (T&E) initiatives such as early integrated testing, cost reduction by sharing high-demand testing assets, and mission-based T&E. The creative and innovative power of the representative set of users employing the software system is continually demonstrated. Though users remain the backbone of every initial operational test (IOT) final exam, this study suggests that the same representative set of users are a potentially viable resource for guiding more efficient early development and testing of future software-intensive systems.

Key words: Integrated testing; mission-based T&E; operational testing; software; users.

Software is incredibly pervasive and evolving at an accelerating pace both within and outside of the military. In fact, according to a recent National Research Council letter report, "software has become essential to all aspects of military system capabilities and operations" (NRC 2008).

The complexity of modern systems creates significant challenges for operational testers. Department of the Army Pamphlet 73-1 states that the operational test (OT) phase focuses on the generation of operational test data under the control of the operational tester with typical user personnel in an appropriate operational environment using production representative systems (U.S. Army 2003). During an OT, there are usually a significant number of data elements required for collection. Sometimes the evaluators provide specific measures related to software for the U.S. Army Operational Test Command (USAOTC) to collect; however, usually the information that USAOTC provides related to software consists of general statements that explain whether the software enabled effective mission task completion when the user employed specific software modules related to the task. Therefore, data on software performance are often provided as a byproduct related to mission-specific data elements.

In this article, operational test events for three separate acquisition programs are examined in chronological order. Two of these test programs are the UH-60M Baseline and UH-60M Upgrade helicopters, and the last program is the MQ–1C Unmanned Aircraft System Extended Range/Multipurpose Quick-Reaction Capability 2 (QRC2). These three programs are ACAT 1D programs and were not specifically classified as ACAT 1A (like many automated information systems or software-intensive systems). In each of these test events, the generalized effect of the software system on the test results is examined. In addition, these results support discussions related to many of the current test and evaluation (T&E) initiatives mentioned throughout several cited articles such as early integrated testing (Streilein and Luna 2009), cost reduction by sharing of high-demand testing assets (Wilson 2009), and mission-based test and evaluation (Streilein 2009; Apicella, Wyant, and Wilcox 2009). Throughout these observations and discussions, the creative and innovative power of the representative set of users employing the software-intensive systems is continually demonstrated. Though users remain the backbone of every initial operational test (IOT) final exam, the results from this study suggest that they are also a frequently untapped but viable resource for guiding more efficient early development and testing of future software-intensive systems.

Test observations and comments

The U.S. Army Operational Test Command (USAOTC), Aviation Test Directorate conducted

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **MAR 2011**	2. REPORT TYPE	3. DATES COVERED **00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE **Operational Testing of Software-Intensive Systems: Observations and Comments**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **U.S. Army Operational Test Command,91012 Station Avenue,Fort Hood,TX,76544-5068**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

the UH–60M IOT Phase I at Fort Hood and Camp Bowie, Texas, October 16–December 8, 2006. This phase included an engineering company, an attack helicopter troop, and a battalion (-) of opposing forces that served as the supporting unit. The test unit consisted of a lift helicopter troop. The test players (16 pilots and 20 crew chiefs/maintainers) participated in missions during the test. Including the pilot test, 37 missions were conducted. There were 113 sorties by individual aircraft that accrued 258.6 hours of flight time (Brown et al. 2007). The IOT Phase I was integrated into the unit's capstone training event when the unit was evaluated against their Mission Essential Task List (METL) tasks by the 21st Cavalry Brigade at Fort Hood, Texas. This approach allowed the tester to have access to many high-demand soldier and equipment assets with reduced costs, while also enabling robust scenarios that supported a simple but effective mission-based T&E approach.

The primary upgrade that the UH-60M Baseline offered, when compared with the UH-60 A/L, was the addition of digital cockpit avionics. The UH-60M Baseline also had the Improved Vehicle Health Monitoring System (IVHMS) on board during the test, which acts similar to a black box recorder for much of the avionics data as well as for many health-monitoring sensors strategically placed across the aircraft. The functionality of the digital cockpit was a major data source for the test, while the IVHMS provided the tester with an avionics data stream that required no additional use of resources by the tester. A byproduct of adding more software to a system often creates opportunities to access the input data used by the software after mission completion. These data are vital when trying to analyze specific test anomalies.

UH-60M aircrew members self-evaluated the capability of the helicopter to complete all missions. According to the aircrews, the UH–60M allowed mission accomplishment with no workarounds for 90% of the IOT Phase I sorties, and an additional 9.5% with workarounds. Aircrew members felt the UH–60M did not allow completion of the mission for 0.5% of the sorties. The bottom line here is that the aircraft was a large step forward when compared with the UH-60A/L. The ability of the pilots to utilize workarounds and redundant systems overshadowed many of the digital cockpit glitches that were attributable to software shortcomings.

At the completion of the UH-60M baseline IOT, the test team compiled the 21 most suggested improvements that were entered by the pilots into the database and presented these improvements back to the pilots in a survey. The pilots were then asked to prioritize the list from most important to least important (1 = most important). *Table 1* shows the results.

It is significant to note that the top eight suggested improvements were all tied to the software/hardware interface for the digital cockpit. Improvement No. 1 (enable separate warm reboot of all processors) became very important to the pilots, because it was discovered during the test that the preflight checklist provided to the pilots was not developed for the situation in which the digital systems on the aircraft were all initializing together off of the Auxiliary Power Unit in an operationally required expedient manner. More specifically, the preflight checklist used to train the pilots before the IOT was developed using a clean, stable, hanger provided, power source in a deliberate and unhurried manner. Therefore, throughout the IOT the pilots continually refined the checklist to perform better for operational missions but regularly embarked with less than a fully functional digital cockpit. If the pilots then encountered a task where a non-mission essential function became essential, they would often choose to land the aircraft and attempt to reboot the whole digital system or simply try to reset specific circuits in an attempt to regain the nonfunctioning capability. The suggested warm reboot would allow them to reset processors without requiring the aircraft to land. It was a workaround to the common problems associated with what was at that time a less than mature software/hardware preflight software initialization process. However, the capability improvement between the new and old system was so significant that the pilots had enhanced mission capability even when the UH-60M digital cockpit was less than fully functional. Because this was an operational test, strict engineering data describing the different failure modes of the interacting interfaces were not captured; however, it is interesting to note that as time progressed, the pilots collectively were able to adjust the preflight checklist in an effort to allow a higher percentage of operational digital subsystems to operate properly.

Improvements Nos. 2–8, which are explained in more detail below, were all requests for additional functionality. The test players were all quite experienced with computers and were well aware that this new cockpit was similar to a computer. Therefore, the pilots made comments throughout the test about functionality that did not already exist and would be most helpful to them for mission success. Improvement No. 2 (enable loading of multiple flight plans) offered an enormous potential time savings as well as the ability to have the system prepared for contingency missions. Improvement No. 3 (the ability of the aircraft to hold all data on shut down) was important because

Table 1. Aviator exit survey rank (sample size = 16) of suggested improvements for UH-60M Baseline Initial Operational Test (IOT) Phase I (ranked by mean).

Suggested improvement	Mean	Median	Standard deviation
1. Enable separate warm reboot of all processors.	2.3	1	2.35
2. Enable loading of multiple flight plans.	4.5	4	3.34
3. Save data on shutdown (Comm/Nav/COMSEC).	5.6	6	4.08
4. Improve radio reliability.	5.6	4	4.29
5. Enable loading of multiple waypoints.	6.8	7	3.33
6. Enable aircraft use of all Aviation Mission Planning System (AMPS) symbols and functionality.	7.6	7	5.66
7. Enable reversal of flight plan.	7.7	8	2.67
8. Enlarge AMPS/Aircraft capability to store local waypoint data.	8.4	9	4.66
9. Make crew chief seats more comfortable with more head clearance.	9.4	12	5.95
10. Improve crew chief restraint system (allow for quicker on/off).	9.6	8	5.95
11. Improve left red position light for night operations.	10.0	11	6.51
12. Add one push target store button.	10.9	11	6.18
13. Eliminate Flight Management System to Multi-Function Display (MFD) transfer delays.	11.2	12	6.18
14. Add GPS vertical navigation capability.	11.3	12	4.86
15. Integrate preflight performance checks into UH-60M digital cockpit.	12.6	11	5.74
16. Improve searchlight.	14.2	17	6.24
17. Eliminate false cautions/advisories.	14.8	17	4.76
18. Enable crew chief visibility to instruments.	15.0	16	5.41
19. Improve Heads Up Display (HUD).	16.0	18	5.78
20. Improve Blue Force Tracker (BFT)/Joint Variable Message Format (JVMF) functionality.	16.3	18	4.92
21. Provide more heat in rear of aircraft.	16.8	19	4.09

Comm/Nav/COMSEC, communication frequencies/navigation information/communication security; GPS, global positioning system.

pilots did not want to reenter data manually each time they landed and shut down an aircraft during a mission; under tight mission timing constraints this required too much time. Furthermore, if required to land to perform a shutdown in an effort to regain the loss of a subsystem, the manual reentry of important parameters became both a loss of time as well as a nuisance. Improvement No. 4 (the ability to improve radio reliability) was also connected to software/hardware issues related to the preflight checklist. Throughout the IOT, nonoperational radio subsystems were a common occurrence due to preflight initialization anomalies, but the aircraft were still considered mission capable because of the redundant design of the radio system. Improvements Nos. 5 and 7 were simple improvements that would result in significant workload reduction for the pilots. Improvements Nos. 6 and 8 were associated with reduced ability of the digital cockpit to efficiently and effectively interface with the Aviation Mission Planning System (AMPS) software that was used by the pilots. Improvements Nos. 12, 14, and 15 are requests for additional software functionality, while No. 13 is related to a sluggish software/hardware process that results in a time loss. Improvements Nos. 17, 19, and 20 are all related to needed improvements of the corresponding software/hardware subsystems. Therefore, 15 of the 21 (71%) suggested improvements were related to software.

Phase II of the UH-60M Baseline IOT consisted of small excursions that were essentially integration tests for the HH-60M (Manning et al. 2007a; medical evacuation version) medical equipment package and the addition of the Common Missile Warning System (CMWS) and Airborne Radio Communications-231 (ARC-231) radio on the baseline aircraft (Manning et al. 2007b). Each of these integration tests consisted of five missions and approximately 10 flight hours. The integration tests were focused only on the data elements corresponding to the newly integrated subsystems. Both of these tests were considered combined Developmental Test/Operational Test (DT/OT) events. Because of the late integration of the systems and the required DT/OT test metrics related to each, single events were planned for each subsystem in an effort to capture information on all developmental and operational issues. For safety reasons, a single Experimental Test Pilot (XP) was used as one of the two pilots in each of the combined DT/OT events.

During the HH-60M excursion, one of the two pilots was an XP who had worked with the UH-60M throughout the DT period and therefore had helped

develop the preflight checklist used during IOT Phase I. Because Medevac missions require a rapid response, while also requiring that the digital cockpit is fully functional in order to safely perform all operations, this test was a major challenge even for the experienced XP. The shortcomings of the preflight checklist not only delayed the rapid response missions but also caused scattered spurious behavior of the digital cockpit systems.

During the UH-60M Baseline CMWS and ARC-231 radio excursion, the pilots consisted of one XP and one line pilot who flew during the UH-60M Baseline IOT Phase I. The scope of this test was limited but required a fully functional digital cockpit in order to properly test the added subsystems. Similar to the HH-60M excursion, the assigned XP had worked with the UH-60M throughout the DT period and therefore had helped develop the existing preflight checklist. Even after the refinement to the checklist provided by the HH-60M XP, it was noted that the line pilot was providing previously undocumented information to the XP on preflight checklist improvements that were learned during the Phase-I IOT. This demonstrated the idea that even DT can benefit from adopting an operational flavor, especially when it comes to developing the aircraft initialization checklist for digital systems. Software and hardware often perform differently as timing and environmental variables change especially when the subsystems interact through the system software. Furthermore, a CW2 line pilot with previous UH-60M experience had comparable abilities to the trained XP when troubleshooting and improving upon the detailed process for properly and quickly performing preflight initialization of the digital cockpit.

The second program of record was the UH-60M Upgrade helicopter program. For this program, the author was involved with a team executing a Limited User Test (LUT) using a flight and software aircraft simulation with actual aircraft subsystems at the Systems Integration Laboratory (SIL) at Huntsville, Alabama, October 6–24, 2008. The primary upgrade to the UH-60M Upgrade aircraft was the incorporation of a full-authority Fly-By-Wire (FBW) flight control system (FCS) into the UH-60M airframe, including advanced, multi-mode flight Control Laws (CLAWS), active cyclic and collective sticks, and integration with the Common Avionics Architecture System (CAAS). Essentially, all existing mechanical control system components upstream of the primary servos were replaced by the FBW FCS. Both the UH-60M Baseline and UH-60M Upgrade program used early user working groups and LUTs focused specifically on cockpit software inside simulated cockpits. Both test programs received benefits from LUTs that provided pre-IOT user feedback on their systems from a small sample set of pilots.

Similar to the IOT Phase I, the pilots found some functionality that the CAAS cockpit lacked; however, for the most part, the suggested improvements made to this cockpit required an even larger leap forward for the software functionality when compared with the suggested improvements from the UH-60M Baseline IOT. The CAAS cockpit had been used operationally before and, therefore, was a more mature system as it benefited from the increased user feedback and development time compared with the UH-60M Baseline digital cockpit. *Table 2* shows the suggested improvements provided by the three aircrews that participated in the LUT; the rating scheme used for this table was different than that used for the UH-60M IOT Phase I. The scheme in *Table 2* focused more on understanding what improvements were necessary before the pilots would consider the UH-60M Upgrade aircraft mission-capable.

It is significant that *Table 2* shows that three of the top five suggested improvements were related to the FBW flight controls. There are many potential explanations. In order to fairly explain these results, it is necessary to describe the key features of the UH-60M FBW flight controls. A key design goal for the UH-60M Upgrade FCS is to provide CLAWS, which will enable Level-1 handling qualities at low speed and in degraded visual environments without compromising the maneuverability of the aircraft throughout the remainder of the mission. Sikorsky translated this requirement into a set of multi-mode, attitude command CLAWS with functions such as low-speed and high-speed turn coordination, and automatic altitude, flight-path, hover, heading, velocity, and position hold modes. Automatic modification of control law response types and modes is handled using regime recognition and task tailoring with the aid of aircraft sensors and pilot vehicle interfaces. The system provides transitions between control modes based on aircraft state and pilot input without the need for the pilot to release the controls to select these modes. The active cyclic and collective enable the use of tactile cueing to provide control mode feedback to the flight crew (Fletcher et al. 2008).

Without going deeper into FBW specifics, the flight handling for this aircraft had a completely new characterization that was not intuitive to experienced UH-60L pilots. As stated in the previous paragraph, the FBW software was designed to change flight control performance and character based upon current flight conditions. This was an entirely new idea for pilots who had used the previous mechanical interfaces

*Table 2. Aviator exit survey ratings (sample size = 6) of suggested improvements for UH-60M Upgrade after Limited User Test (LUT).**

Proposed improvement	Mean	Min	Max
1. Eliminate frequent flight control unlinked scenarios.	4.8	4	5
2. Fix HUD problems.	4.0	3	5
3. Lower flight controller conflict audio volume.	4.0	1	5
4. Improve flight control performance during takeoffs and landings.	4.0	1	5
5. Improve flight control performance during high to low and low to high aircraft speed transitions.	3.8	1	5
6. Integrate flight director.	3.8	3	5
7. Create JVMF indication that does not time out on the MFD.	3.8	3	5
8. Decrease size of HUD image.	3.7	1	5
9. Correct AMPS transfer shortcomings (airspeeds, altitudes, and times for route do not transfer).	3.5	0	5
10. Provide a Warning or Caution when Aircraft Survivability Equipment (ASE) fails.	3.3	1	5
11. Use the same Navigation symbols on MFD displays as are used for AMPS system.	3.3	3	5
12. Create separate map control channels for pilot and copilot.	3.2	3	4
13. Provide more than 99 waypoints.	3.2	0	5
14. On the Central Display Unit (CDU), provide a duplicating feature, so when performing flight plan management you don't have to change each point to update the route, altitude, or airspeed.	3.2	2	5
15. Create CLEAR button on CDU that will clear last character on scratch pad when pushed, and pressing and holding CLEAR should clear the entire field.	3.2	2	4
16. Create return or back button on all CDU pages and locate it in the same place.	3.0	1	4
17. Have remote radio select cycle through active radios only.	3.0	2	4
18. Increase map storage capability to 100+ GB.	3.0	0	5
19. Provide QWERTY-style keyboard for JVMF messaging.	3.0	1	5
20. Make collective trim beeper faster or rate adjustable.	2.8	2	3
21. Improve map displays—too much white. Active leg and aircraft should be different color.	2.8	0	5
22. Change flight controller conflict audio to a less obnoxious sound.	2.7	0	5
23. Ensure that engine out meets two gates instead of just gas generator speed (NG).	2.7	1	4
24. On System Page, when something is wrong with an item, denote with a red X instead of a checkmark.	2.7	1	3
25. Fix select feature on Multi-Function Slew Controller (MFSC) for JVMF.	2.7	1	3
26. Modify fuel page so that average fuel burn is a combined number for both engines.	2.7	1	3
27. Make ASE status available on primary screens.	2.5	1	4
28. Improve delay related to collective radio toggle switch.	2.5	1	4
29. Make Compass rose on MFD maps have adjustable colors for better visibility.	2.3	1	3
30. Include more details with Check Aircraft Status advisory.	2.3	1	3
31. On fuel page, display a 15-minute moving average combined fuel flow.	2.3	1	4
32. Add ability to see current cursor position with MFSC.	2.3	0	5
33. Provide the ability to pull against the collective detent without coming out of it in forward flight and to establish a vertical speed greater than the current threshold of 400 fpm (~700 fpm).	2.3	0	5
34. On fuel page, add the ability to access instantaneous fuel burn for both engines together.	2.3	1	3
35. Add AHEAD/BEHIND display to mission page on the CDU.	2.0	1	4
36. Change color of MFSC icon.	1.8	0	3
37. Place Turbine Gas Temperature on Vertical Situation Display (VSD).	1.8	0	4
38. Show Hover point on Horizontal Situation Display Hover page as soon as you are on approach to that point.	1.8	1	3
39. Create ability to display air control point list.	1.8	0	4
40. Add the question mark "?" to the JVMF keyboard.	1.7	1	3
41. Put all equipment status information on one page.	1.5	0	3
42. Provide fixed lubber line on the Horizontal Situation Indicator under the HEADING display window.	1.5	0	3
43. Make system default to Horizontal Situation Display Hover page.	1.2	0	3

Min, minimum; Max, maximum; HUD, heads-up display; JVMF, Joint Variable Method Format; MFD, multi-functional display; AMPS, Aviation Mission Planning System; ASE, aircraft survivability equipment.

*Rating Scale: 5 = Change required; don't want this aircraft without the change; 4 = Change must happen quickly, but can execute mission until fixed; 3 = Change extremely helpful; significant improvement; really want this to happen but would fly without; 2 = Change has value but can definitely execute mission without; low priority; 1 = Change helps; small improvement; good idea but won't help me much; 0 = Change not required; won't help me at all.

that only changed character in response to the physics of the current situation. Sikorsky chose a customizable software solution to meet the requirement provided to them. The challenge then was to train the previously trained UH-60L pilots to use the newly designed software-driven flight controls effectively.

During the LUT, the pilots successfully performed the missions assigned to them in the simulator; however, there was one key negative finding. All of the pilots, on at least one occasion during the test, stated that they were fighting the controls. Those statements usually occurred in reference to terrain

flight while either evading threats or avoiding inadvertent instrument meteorological conditions. Basically, when placed in a stressful situation, almost all of the pilots reverted back to flying the aircraft similar to the way they were trained to fly a UH-60L in tight tactical situations. The fighting of the controls was systematic and was a simple power struggle between the pilots and the software build that was current at that moment. For the final two aircrews, USAOTC was able to collect and provide a general analysis of Sikorsky flight control electronic files that demonstrates this phenomenon. These files and many other issues are described in more detail in the USAOTC Abbreviated Operational Test Report (AOTR) (Pontes and Manning 2008).

The last program described here is the MQ–1C Unmanned Aircraft System Extended Range/Multi-purpose QRC 2 LUT that was conducted at Edwards Air Force Base (AFB), California, and the National Training Center (NTC), Fort Irwin, California, from May 19 through June 4, 2010, during an NTC rotation. During the test, the MQ–1C QRC2 unit conducted reconnaissance, surveillance, security, target acquisition, attack, battle damage assessment, and communications relay missions over 179 flight hours (Brown et al. 2010). It is interesting to note that the test team collected and processed more than 17 terabytes of video and telemetry data to describe those 179 flight hours. This was a significantly larger data output when compared with the previously described helicopter tests.

During the MQ-1C QRC2 LUT, software/hardware shortcomings manifested themselves in the area of suitability. There were numerous Test Incident Reports (TIRs) attributable either to One–System Ground Control Station (OSGCS) software or hardware instability. The common workaround to these issues was to simply reboot or refresh the system or subsystem that displayed severely degraded performance. The success of this workaround was variable and sometimes entirely unsuccessful. The instability of the OSGCS resulted in numerous Reliability, Availability, and Maintainability (RAM) system aborts based on the failure definition/ scoring criteria intact during the LUT. It was determined by the Program Manager (PM) after the test that this instability was related to both the software present as well as the heavy load on the computer processors during the test.

Because this system was a QRC that required significant additional functionality, it is understood that both the software and hardware were performing at the limits of current capability. Additionally, because of the requirement for software to both fly the aircraft and provide the capability for the operators to manage all of the subsystems, there is also the challenge of software task prioritization. A complete list of all of the system strengths and weaknesses is presented in detail in the MQ-1C QRC2 LUT AOTR (Brown et al. 2010).

Similar to the previously described tests, 12 of the 19 (63%) system weaknesses and three of the top four mentioned in the MQ-1C Test Report were attributable to software/hardware performance or design shortcomings.

Unlike the two helicopter programs mentioned previously, the MQ-1C program had QRC objectives; therefore, it was exposed to operational testing earlier in the system development process. Even though early operational testing does not guarantee success in the final exam IOT, currently scheduled for the summer of 2011, it certainly improves the PM's potential probability for success. The MQ-1C QRC2 LUT results have focused the PM and the test community directly on the most important operational shortcomings related to mission success with approximately 1 year available to employ the DT test-fix-test process to the user prioritized list of shortcomings.

Discussion

Software-intensive systems require user and system inputs for effective operations. For each of the systems discussed in this article, the test team was able to capture data streams that were already inherent to the system in an effort to more effectively describe strengths and weaknesses in performance. It is these data streams that often provide the ability to explain unanticipated processes and events. Operational testers must make a dedicated effort to use this type of data that is captured on an operationally noninterference basis instead of requiring the resources and paperwork associated with connecting test-specific data-collection hardware. It requires early planning and work to enable the proper collection of these data streams, but they have proved invaluable for answering questions in each test described in this article.

Because the author's directly acquired operational testing knowledge spans only the tests listed in this article, it is not possible to offer strong comparisons between the testing described here and the testing conducted during time periods when systems were less software intensive. However, it is argued that as systems have become more software intensive, military users have become more creative and demanding. Software and the processors that enable it are so heavily embedded into everyday life functions that soldiers are qualified to offer powerful opinions on shortcomings and functionality multipliers for software-intensive systems. Their everyday experiences have taught them how both good and bad software

behave. The inherent ability of the operational soldier to effectively and efficiently shape software-intensive systems provides a rapid prototyping toolkit that offers acquisition programs more effectiveness and efficiency. The payoff for discovering software redesigns up-front and early in a program are huge. As shown in the observations from the tests described in this article, if provided an opportunity, a soldier always has ideas on how to improve a software-intensive system in order to enhance mission success.

The bundling of several operational tests on similar test schedules has been suggested to the test community as a viable effort to more efficiently use resources. The primary risk factor associated with employing this method is the inherent schedule risk associated with those programs entering Initial Operational Test & Evaluation (IOT&E). Early user test bundling initiatives offer similar efficiency rewards with less program schedule risk; early feedback to PMs using prototypes or simulators adds more value and certainty to the ongoing development efforts. Robust operational assessment information from users even when provided in smaller sample sizes has proven invaluable during the many ATEC Forward Operational Assessment (FOA) deployments.

Mission-based T&E and design of experiments are two more initiatives that testers must currently consider. Of interest, both the UH-60M Baseline IOT Phase I and the MQ-1C QRC2 LUT occurred during training exercises. This occurrence provided the tester increased access to high-demand assets (soldiers and support equipment) that enabled more operationally relevant test scenarios. Though some control of the test event is sacrificed in this type of setup owing to the inability to enforce strict experimental design, it is possible to use instrumentation and documentation to extract DT-type metrics after completion of each test mission. Additionally, during these exercises, the soldiers' focus is always weighted to mission success, which enables high-quality OT data output.

Conclusion

The objective of the many T&E initiatives mentioned here is to more efficiently and effectively facilitate the fielding of equipment to warfighters. Users have noted past deficiencies in defense acquisition and testing by explaining that they were simply provided equipment and told to make it work. The current test process for most systems provides equipment to a representative group of users who are empowered to provide feedback on the systems at the end of the system development cycle. Based on the test results discussed in this article, there is a good argument for continuing the final exam IOT process while also using early user tests as an experimental guide to focus the development effort on the most important mission-enabling areas (Crevecoeur 2010). Early user feedback on software-intensive systems is an efficiency and effectiveness multiplier. Culturally, the test and acquisition community must embrace the fact that "test failures, especially during early development should be received as learning opportunities and chances to solve problems as they are uncovered" (DOD 2007). Following the same idea, another recent ITEA publication argued that T&E needs to evolve, so that it is directly integrated into the system development process (Weiss, Roberts, and Cross 2009).

There are obvious limitations to testing software systems during the IOT phase. Operational test teams focus more on enabling evaluators to relate test results to actual operations than providing the engineering data required to allow them to understand the specific failure of a software module. Intensive software testing is rightly performed during DT. However, it is a mistake not to allow groups of operational soldiers to steer the development of software systems early in the acquisition program development. PMs may consider early operational testing as too risky to program success especially for programs with Office of the Secretary of Defense (OSD), Director, Operational Test and Evaluation (DOT&E) oversight. Acquisition decision makers must consider a cultural change that rewards early OT and does not punish PMs who choose early OT as a tool to more efficiently and effectively guide the development of their system. It is a valid argument that more operationally realistic DT provides similar information as early OT minus the larger, more representative, set of users. However, the observations from the tests described here demonstrate the creative and evaluation capabilities of a representative set of users. The representative set of users is the backbone of every IOT; however, they remain mostly an untapped resource for guiding the important early development and testing of the growing number of software-intensive systems. ❑

WILL MANNING is an operation research analyst with the Aviation Test Directorate at the U.S. Army Operational Test Command, Fort Hood, Texas. Previously, he worked as a physical scientist at U.S. Army Research, Development, and Engineering Command (RDECOM) Communications-Electronics Research, Development, and Engineering Command (CERDEC) Night Vision and Electronic Sensors Directorate, Fort Belvoir, Virginia. Before working for the Army, he was employed as a research scientist studying passive millimeter-wave and microwave radiometry at the Joint Center for Earth Systems Technology at the

University of Maryland, Baltimore County in support of the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center. He received his bachelor and master of science degrees in meteorology from Texas A&M University at College Station, Texas. He has received Defense Acquisition University (DAU) Level-3 certifications in test and evaluation, systems engineering, and program systems engineering. E-mail: will.manning@us.army.mil

References

Apicella, F. J., K. W. Wyant, and C. Wilcox. 2009. ATEC initiatives in response to the office of the secretary of defense policy guidelines for test and evaluation. *ITEA Journal* 30 (3): 361–368.

Brown, D., E. Larson, W. Manning, J. Moltenberry, and C. Anderson. 2010. "Abbreviated Operational Test Report for MQ-1C Unmanned Aircraft System Extended Range/Multipurpose Quick-Reaction Capability 2 Limited User Test," July 2010. Fort Hood, TX: U.S. Army Operational Test Command.

Brown, D., W. Manning, J. B. Bush, and J. Kennedy. 2007. Test report for Black Hawk Utility Helicopter (UH-60M) Initial Operational Test Phase I. *U.S. Army Operational Test Command*, February 2007.

Crevecoeur, R. 2010. Changing norms: Keeping Army test and evaluation flexible and adaptable. *ITEA Journal* 31 (3): 393–398.

Deputy Under Secretary of Defense (Acquisition, Technology, and Logistics). 2007. Department of Defense report to Congress on policies and practices for test and evaluation, p. 6. Washington, D.C.: Department of Defense.

Fletcher, et al. 2008. UH-60M Upgrade fly-by-wire flight control risk reduction using the RASCAL JUH-60A in-flight simulator. In *Proceedings of the American Helicopter Society 64th Annual Forum*, April 29–May 1, Montreal, Canada. http://www.vtol.org/f64_bestPapers/handlingQualities.pdf (accessed December 15, 2010).

Manning, W., J. B. Bush, C. Scott, and C. Arnold. 2007a. Test data report for the Black Hawk Utility Helicopter (UH-60M) Initial Operational Test Phase IIa, Medical Evacuation Helicopter (HH-60M) Excursion. *U.S. Army Operational Test Command*, February 2007.

Manning, W., J. B. Bush, C. Scott, and C. Arnold. 2007b. Test data report for the Black Hawk Utility Helicopter (UH-60M) Initial Operational Test Phase IIb, Common Missile Warning System/Airborne Communications Radio – 231 (CMWS/ARC-231) Excursion. *U.S. Army Operational Test Command*, April 2007.

NRC. 2008. Preliminary observations on DoD software research needs and priorities: A letter report. Committee on Advancing Software-Intensive Systems Producibility, National Research Council. Washington, D.C.: National Academy of Sciences.

Pontes, J. W., and W. Manning. 2008. Abbreviated operational test report for Black Hawk Utility Helicopter (UH-60M Upgrade) Limited User Test, Fort Hood, Texas. *U.S. Army Operational Test Command*, December 2008.

Streilein, J. J. 2009. Test and evaluation of highly complex systems. *ITEA Journal* 30 (1): 3–6.

Streilein, J. J., and N. D. Luna. 2009. ATEC's approach to integrated testing. *ITEA Journal* 30 (3): 355–359.

U.S. Army. 2003. Pamphlet 73-1. *Test and evaluation in support of systems acquisition.* 30 May 2003. Washington, D.C.: Department of the Army Headquarters, http://www.apd.army.mil/series_range_pubs.asp (accessed December 15, 2010).

Weiss, L. G., R. Roberts, and S. E. Cross. 2009. The Systems Engineering and Test (TSET) approach for unprecedented systems. *ITEA Journal* 30 (3): 386–394.

Wilson, B. 2009. Integrated testing: A necessity, not just an option. *ITEA Journal* 30 (3): 375–380.